York University
Human Participants Review Committee
Guideline 5: Secondary Use of Data

1. Data Collection, Use and Security: What are Researcher's Responsibilities?

Researchers are required to safeguard the information and data collected via the use of human participants in their research. As stated in the TCPS (2014), "there is widespread agreement about the interests of participants in protection of privacy, and the corresponding duties of researchers to treat personal information in a confidential manner. Indeed, the respect for privacy in research is an internationally recognized norm and ethical standard. "

As it speaks to the collection and use of data within the context of responsibilities of researchers to their research participants, there are three main considerations: privacy, confidentiality and security.

> **Privacy:** Refers to an individual's right to be free from intrusion or interference by others. An important aspect of privacy is the right to control information about oneself. The concept of consent is related to the right to privacy. Privacy is respected if an individual has an opportunity to exercise control over personal information by consenting to, or withholding consent for, the collection, use and/or disclosure of information.
>
> **Confidentiality:** The ethical duty of confidentiality refers to the obligation of an individual or organization to safeguard entrusted information. The ethical duty of confidentiality includes obligations to protect information from unauthorized access, use, disclosure, modification, loss or theft. Fulfilling the ethical duty of confidentiality is essential to the trust relationship between researcher and participant, and to the integrity of the research project.
>
> **Security:** Security refers to measures used to protect information. It includes physical, administrative and technical safeguards. An individual or organization fulfils its confidentiality duties, in part, by adopting and enforcing appropriate security measures. Physical safeguards include the use of locked filing cabinets, and the location of computers containing research data away from public areas. Administrative safeguards include the development and enforcement of organizational rules about who has access to personal information about participants. Technical safeguards include use of computer passwords, firewalls, anti-virus software, encryption and other measures that protect data from unauthorized access, loss or modification.

Ethical concerns regarding privacy decrease as it becomes more difficult (or impossible) to associate information with a particular individual. These concerns also vary with the sensitivity of the information and the extent to which access, use or disclosure may harm an individual or group.

The easiest way to protect participants is through the collection and use of anonymous or anonymized data, although this is not always possible or desirable. For example, after information is anonymized, it is not possible to link new information to individuals within a dataset, or to return results to participants. A "next best" alternative is to use de-identified data: the data are provided to the researcher in de-identified form and the existing key code is accessible only to a custodian or trusted third party who is independent of the researcher. The last alternative is for researchers to collect data in identifiable form and take measures to de-identify the data as soon as possible. Although these measures are effective ways to protect

participants from identification, the use of indirectly identifying, coded, anonymized or anonymous information for research may still present risks of re-identification.

Technological developments have increased the ability to access, store and analyze large volumes of data. These activities may heighten risks of re-identification, such as when researchers link datasets or where a dataset contains information about a population in a small geographical area, or about individuals with unique characteristics (e.g., uncommon field of occupational specialization, diagnosis of a very rare disease). Various factors can affect the risks of re-identification, and researchers and the Ethics Review Committee alike are advised to be vigilant in their efforts to recognize and reduce these risks. Data linkage of two or more datasets of anonymous information may present risks of identification.

Where it is not feasible to use anonymous or anonymized data for research (and there are many reasons why data may need to be gathered and retained in an identifiable form), the ethical duty of confidentiality and the use of appropriate measures to safeguard information become paramount.

Therefore, researchers must be cognizant of the potential implications of a breach of privacy of their research participants' data. Privacy risks in research relate to the identifiability of participants, and the potential harms they, or groups to which they belong, may experience from the collection, use and disclosure of personal information. Privacy risks arise at all stages of the research life cycle, including initial collection of information, use and analysis to address research questions, dissemination of findings, storage and retention of information, and disposal of records or devices on which information is stored.

Breaches may also occur through the use of secondary data analysis.

2. What is secondary data analysis?

The TCPS[1] defines secondary use of data as follows:

> Secondary use refers to the use in research of information originally collected for a purpose other than the current research purpose. Common examples are social science or health survey datasets that are collected for specific research or statistical purposes, but then re-used to answer other research questions. Information initially collected for program evaluation may be useful for subsequent research. Other examples include health care records, school records, biological specimens, vital statistics registries or unemployment records, all of which are originally created or collected for therapeutic, educational or administrative purposes, but which may be sought later for use in research. Chapter 12 provides further guidance on research involving secondary use of previously collected biological materials.
>
> Privacy concerns and questions about the need to seek consent arise, however, when information provided for secondary use in research can be linked to individuals, and when the possibility exists that individuals can be identified in published reports, or through data linkage (Article 5.7). Privacy legislation recognizes these concerns and permits secondary use of identifiable information under certain circumstances

[1] TCPS 2 (2014)

What is "Identifiable" Information?"

Where researchers seek to collect, use, share and access different types of information or data about participants, they are expected to determine whether the information or data proposed in research may reasonably be expected to identify an individual. For the purposes of this Policy, researchers and Research Ethics Boards (REBs) shall consider whether information is identifiable or non-identifiable. Information is identifiable if it may reasonably be expected to identify an individual, when used alone or combined with other available information. Information is non-identifiable if it does not identify an individual, for all practical purposes, when used alone or combined with other available information. The term "personal information" generally denotes identifiable information about an individual. The assessment of whether information is identifiable is made in the context of a specific research project.

Anonymous vs Anonymized Data: Ethics Review & Consent Requirements

There is a distinction to be made between anonymous data and anonymized data. As per the TCPS2, anonymous data is data for which the information never had identifiers associated with it (e.g., anonymous surveys) and risk of identification of individuals is low or very low. Identifiers are information that could potentially identify an individual; alone or in combination with other information that was collected.[2]

For information to be anonymous no direct or indirect identifiers were ever collected therefore the "risk of identification of individuals is low or very low". This is different than anonymized data in which the information or materials have been "irrevocably stripped of direct or indirect identifiers, a code is not kept to allow future re-linkage, and risk of re-identification of individuals from remaining identifiers is low or very low".

Research that relies exclusively on secondary use of anonymous information, or anonymous human biological materials, may not require ethics review so long as there is no process of data linkage and the recording or dissemination of results does not generate identifiable information

Researchers are required, however, to provide the provenance of the data set (ie the source of the data such as a data repository; data collected under an approved protocol etc.) to the office of Research Ethics by completing the "Secondary Data Analysis- Anonymous Data Reporting Form". However, for research involving secondary data analysis of anonymized data, where there is a possibility of data linkage and/or threat to the confidentiality and anonymity of said data, ethics review will be required.

Secondary Use of human Biological Materials

Secondary use refers to the use in research of human biological materials originally collected for a purpose other than the current research purpose. A researcher may seek to use human biological materials left over from a diagnostic examination or surgical procedure, or materials that were collected for an earlier project. Reasons to conduct secondary analyses include: avoidance of duplication in primary collection and the associated reduction of burdens on participants; corroboration or criticism of the conclusions of the original research; comparison of change in a research sample over time; application of new tests of hypotheses that were not available at the time of original collection; and confirmation that the data or materials are

[2] University of Waterloo, "Does my data collection activity require ethics review"

authentic. Privacy concerns and questions about the need to seek consent arise, however, when human biological materials provided for secondary use in research can be linked to individuals, and when the possibility exists that individuals can be identified in published reports, or through linkage of human biological materials with other data.[3]

Researchers, therefore, are not required to seek consent from individuals for the secondary use of non-identifiable human biological materials.

Given the likelihood of data linkage, most genetic material is not considered to be anonymous data and therefore its use may be subject to ethics review. However, making this determination can be challenging due to many factors. Should your research involve human biological materials, please contact the Office of Research Ethics (ore@yorku.ca) for further information and guidance.

In the case of the secondary use of identifiable human biological materials, researchers must obtain consent in accordance with applicable laws, unless the researcher satisfies all the requirements as outlined in Article 12.3, TCPS2nd edition. [4]

Note: Secondary use of human biological materials identifiable as originating from a specific Aboriginal community, or a segment of the Aboriginal community at large, is subject to ethics review. Please consult the Aboriginal Research ethics guidelines for further information.

When is ethics review required?

As stated previously, in general, for research that relies exclusively on secondary use of anonymous information, or anonymous human biological materials, ('anonymous' specifically meaning that the data never had identifiers associated with it at any point in time) and so long as the process of data linkage or recording or dissemination of results does not generate identifiable information ethics review is not required. However, if the data is "anonymous" researchers are asked to consider the following additional questions:

- Where did you get the data? From a previously approved protocol? Regulated data Repository or Archive?
- Are there any restrictions on use or dissemination of findings?
- Any potential for data linkage?
- Is it Aboriginal data?

Note: Researchers are advised that for research involving secondary analysis of anonymous data, Researchers are required, to complete the "Secondary Data Analysis- Anonymous Data Reporting Form" and file with the Office of Research Ethics.

[3] TCPS 2nd edition, chapter 12
[4] ibid